Abstract

A DNA Sequence of a Bacterium generally consists of a sequence of four different nucleotides represented by English alphabets A, C, G and T. Most of the genomic regions in a bacterial genome exhibit uniformity in the context of (G+C)%. However, in some of the regions (G+C)% proportion is found to be significantly high whereas in some regions it is significantly low which are called as high and low altered GC regions, respectively. Discovery of several important features in bacteria like horizontal gene transfer, pathogenesis etc., located in the altered GC regions. This gave impetus to the study of altered GC regions in genomes from evolution point of view. We have developed a method based on Hidden Markov Model to identify the high and low altered GC regions in Bacterial Genome. Our method finds altered GC regions of E. coli up to 88.9% accuracy. Our method also predicts a few potentially altered GC regions in E. coli genome.

***Key words**: Altered GC Regions, Markov Chain, Hidden Markov Model*